

# Wrangling report

## Gathering Data for this Project

In this project, I split gathering process to two parts:

**Part 1:** I invested some time to extract the data from the json file given to me `tweet_json`, and storing this data in Data Frame, it contains from some dictionaries, so I read each dictionary and add the useful information in some lists 'each represent a piece of information', then go to the next dictionary and get the data from it and add it to the lists, then I combine all lists to form a final Data Frame which called "df\_gathered".

**Part 2:** I read the csv file `twitter-archive-enhanced.csv` which was given to me using pandas library, and store it in Data Frame which called "df", to go to the next step which is Assess the data.

I get also the image prediction file using Requests library from the URL given in the project description and store it as `image-predictions.tsv`.

## Assessing Data for this Project

In this step I have to search my Data Frames to get all possible issues that I can found, the issues could be quality issues or tidiness issues, then documented all the issues to be ready for the next step 'Clean the data', to fix those issues.

## Cleaning Data for this Project

In this step I have to fix all the issues found in the Assessing step, and test my code after fixing each issue.

Then, I can decide if go to the next step or I found some more issues that need to go to step 2 to evaluate it again.

## Storing Data for this Project

Store the clean Data Frame(s) in CSV form, here I had three Data Frames one is representing the tweets information, and another Data frame has the dog's information like the rating and the name of each dog, and the last one is the image prediction Data Frame, I stored them in one csv file files and called this file `twitter_archive_master.csv`